



JONATHAN WILLIAM WELBURN, AARON STRONG, FLORENTINE  
ELOUNDU NEKOUL, JUSTIN GRANA, KRYSTYNA MARCINEK,  
OSONDE A. OSOBA, NIRABH KOIRALA, CLAUDE MESSAN SETODJI

# Systemic Risk in the Broad Economy

---

Interfirm Networks and Shocks in the  
U.S. Economy



For more information on this publication, visit [www.rand.org/t/RR4185](http://www.rand.org/t/RR4185)

**Library of Congress Cataloging-in-Publication Data** is available for this publication.

ISBN: 978-1-9774-0390-2

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2020 RAND Corporation

**RAND®** is a registered trademark.

*Cover image: Jonathan William Welburn*

### Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit [www.rand.org/pubs/permissions](http://www.rand.org/pubs/permissions).

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

### Support RAND

Make a tax-deductible charitable contribution at  
[www.rand.org/giving/contribute](http://www.rand.org/giving/contribute)

[www.rand.org](http://www.rand.org)

## Preface

---

Following the 2008 global financial crisis, research and policy focused on the risk that heavily interconnected systems could fuel the spread of economic crises across sectors, regions, and countries—a problem that has become known as *systemic risk*. Academic researchers shifted their attention to understanding systemic risk within financial networks while policymakers offered new approaches for promoting financial stability and preventing financial firms from becoming either too big or too interconnected to fail. However, discussions of systemic risk outside the financial sector have been limited. In this report, we address this gap and examine systemic risk in the broad economy. This report describes (1) efforts to gather large data sets on observed inter-firm linkages, (2) procedures to infer missing data on network linkages, and (3) an approach to calibrate likely connections to financial data. Our methodology results in a better estimation of firm-level networks, useful for studying the potential systemic risk posed by firms across sectors in the U.S. economy. We found evidence that systemic risk extends beyond the financial sector to firms in diverse sectors of the economy. Through this methodology, we contribute to the literature on the microfoundations of aggregate economic risk. Furthermore, through these findings, we contribute to the policy discussion on managing aggregate shocks while highlighting the policy need to look beyond the financial sector to areas of systemic risk across the broad economy.

## **RAND Education and Labor**

This study was undertaken by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking. Questions about this report should be directed to [jwelburn@rand.org](mailto:jwelburn@rand.org). For more information, visit [www.rand.org/education-and-labor](http://www.rand.org/education-and-labor), and questions about RAND Education and Labor should be directed to [educationandlabor@rand.org](mailto:educationandlabor@rand.org).

## **Funding**

Funding for this research was provided by gifts from RAND supporters and income from operations.

# Contents

---

**Preface** ..... iii

**Figures and Tables**..... vii

**Summary**..... ix

**Acknowledgments** ..... xiii

**Abbreviations**..... xv

**CHAPTER ONE**

**Introduction** ..... 1

**CHAPTER TWO**

**Literature Review**..... 5

Systemic Risk and Contagion Is Not a New Idea ..... 5

Research Is Increasingly Exploring the Role of Networks in Financial  
Risk ..... 6

The Study of Risk Is Expanding to Include Production Networks ..... 6

**CHAPTER THREE**

**The Model of Firm-Level Linkages** ..... 9

**CHAPTER FOUR**

**Observed Interfirm Production Network**..... 11

Data Sources and Techniques ..... 11

Description of Observed Network ..... 13

Analysis of Observed Network Structure ..... 16

CHAPTER FIVE

**Estimating Production Networks Through Inference**..... 23

Inference for a Complete Network..... 23

Network Calibration..... 26

CHAPTER SIX

**Contagion Across Firm Networks** .....29

CHAPTER SEVEN

**Conclusions and Implications**.....35

Policy Implications..... 36

Limitations and Future Work ..... 38

**References** .....41

# Figures and Tables

---

## Figures

- 4.1. Distribution of Data Coverage Across Sectors ..... 14
- 4.2. Observed Production Network Graph ..... 15
- 4.3. Node Degree Distribution ..... 17
- 4.4. Node Indegree and Outdegree Distribution ..... 18
- 4.5. Network Distance, Shortest Path, and Diameter ..... 20
- 4.6. Distribution of Geodesic Distances in Network **D** ..... 21
- 5.1. Probable Connections Uncovered by Logistic Regression ..... 25
- 6.1. Distribution of Aggregate Losses from Idiosyncratic Shocks.. 30

## Tables

- 4.1. Summary of Node Degree Statistics..... 16
- 5.1. Results of Optimization Model Runs ..... 28
- 6.1. Top 20 Firms with the Highest Estimated Total Network  
Losses..... 32
- 6.2. Top 20 Firms with Highest Estimated Losses Per Unit  
Revenue ..... 33





## Summary

---

Following the economic crisis of 2008, International Monetary Fund chief economist Olivier Blanchard stated that “[w]e have entered a brave new world. The economic crisis has put into question many of our beliefs. We have to accept the intellectual challenge” (Blanchard et al., 2012). In the decade that followed, academic researchers and policymakers shifted their attention to address the intellectual challenge and find new approaches to the measurement and management of aggregate risks. The ability of small, seemingly isolated risks to grow and spread across heavily interconnected systems—a problem summarized by the term *systemic risk*—has emerged as a central focus for research and policy change. A body of research has developed around the study of economic networks, with a particular focus on interbank networks and their relation to economic stability. In addition to research, the question of whether financial institutions were too big or too interconnected to fail has driven policy changes. The 2010 passage of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Pub. L. 111-203, 2010) was a broad response addressing the challenges of the financial crisis with new regulation. However, despite the increased attention on systemic risks, surprisingly little attention has spilled over beyond financial networks into systemic risks in the broad economy.

A primary reason that systemic risk in the broad economy has been understudied might be a lack of data on firm-level networks. That is, although notable research efforts have been made to examine the structure of interbank networks and their contributions to systemic risk or the sector-level origins of aggregate shocks, an inability to directly mea-

sure the interconnections between firms might have kept researchers from understanding the microfoundations of aggregate risk at the level of firms across all sectors of the economy. As a result, a broad exploration of the systemic risk posed by firms outside the financial sector has been largely absent from both academic and policy communities.

This report addresses an important gap in the literature by improving the understanding of systemic risk at the level of individual firms by estimating interfirm production networks in the U.S. economy. We leverage a key U.S. regulation—the Statement of Financial Accounting Standards (SFAS) No. 131, which requires firms to report key customers and suppliers accounting for more than 10 percent of their total sales—to construct a data set of observed network connections between supplier and customer firms. This data set’s inherent observation biases prevent it from straightforward application to firm-level analysis. In fact, the data set constructed solely from SFAS No. 131 observations is missing data in two key dimensions: Linkages falling below the 10-percent threshold are unobserved, and observed linkages omit the flows of goods and services across each connection. We specifically address this problem, analyzing the observed network while using inference techniques to address the challenges of missing data.

That analysis starts with the structure of the observed interfirm network. We reveal a sparse network of supplier-customer linkages with a dense center composed of heavily interconnected firms with numerous customer and supplier linkages. Using network analysis, we explored the distribution of interconnectivity across firms, finding that of the most heavily interconnected firms, many of the most heavily interconnected firms come from several different sectors of the economy. In particular, in addition to financial sector firms, we found many firms in technology and telecommunications with high levels of interconnectivity in observed firm networks. However, to understand the role of systemic risk at the level of firms, our analysis extends beyond the observed firm network.

In the style of Acemoglu, Ozdaglar and Tahbaz-Salehi (2015), Barrot and Sauvagnat (2016), Carvalho and Tahbaz-Salehi (2019), and others, we constructed a firm-level model of production networks for use in input-output analysis. Although the true interfirm network is

unknown, both in magnitude and direction of linkages between firms, we propose a two-step procedure for estimating the true network structure. First, we used statistical inference through logistic regression to estimate the likelihood of unobserved network connections. Second, we calibrated a likely network configuration to empirical data on the inputs and outputs of individual firms using a computational optimization procedure. Using an estimated network configuration, we further estimated the potential aggregate impact of an isolated shock on each individual firm, exploring the systemic risk of individual firms.

We found reason for a renewed look at systemic risk by academic researchers and policymakers alike. It is not just that systemic risk seems to be present outside the financial network (and thereby away from areas of academic focus in the ten years following the 2008 crisis) but also that the firms that are systemically important might have changed since then. Specifically, growth in the technology sector over the past decade has contributed to new forms of systemic risk. No firm epitomizes the shift in systemic risk more than Amazon and its increasingly widespread cloud computing service, Amazon Web Services (AWS), a point illustrated through the efforts of this report. Amazon's centrality in traditional production networks was just emerging at the time of the 2008 crisis. Now, its centrality in digital networks underpinning diverse firms and even public institutions provides an example of the potential of systemic risk in the broad economy, an example that calls for further study on potential risks.

Above all, this report presents an important methodological advancement for the study of the microfoundations of aggregate risk. The applicability of this advancement extends beyond the general study of systemic risk and aggregate shocks and into the study of specific events. For example, firm-level analysis could heighten the understanding of the potential aggregate impact of localized events, such as natural disasters. The estimation of interfirm production networks in this report are a first step to true firm-level analysis. Future research should expand on these methods to provide increasingly accurate estimation of interfirm networks. The result can extend the familiar approaches of sector-level input-output analysis to the level of firms.



## Acknowledgments

---

We would like to thank Craig Bond and Joost Santos for their thorough reviews of this report and their helpful suggestions. We would also like to thank Howard Shatz, Darleen Opfer, Susan Marquis, and many other RAND Corporation colleagues for their support and useful feedback through various stages of this report. Additionally, we thank Sydne Newberry for her helpful feedback and careful edits.



# Abbreviations

---

AWS	Amazon Web Services
COGSS	costs of goods and services sold
EDGAR	Electronic Data Gathering, Analysis, and Retrieval
IT	information technology
SEC	Securities and Exchange Commission
SFAS	statement of financial accounting standards
SIC	Standard Industry Classification
S&P	Standard and Poor's





## Introduction

---

The 2008 global financial crisis altered economists' thinking on macroeconomic risk (e.g., Blanchard et al. 2012). A decade of study later, the academic literature on macroeconomics has expanded with a renewed focus on the key drivers of systemic risk within interbank networks. The focus of policymakers has mirrored the academic literature, focusing on whether some banks were too big and too interconnected to fail. That focal point led to the 2010 passage of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Pub. L. 111-203, 2010) and the creation of the Financial Stability Oversight Council.

In many ways, the policy focus on the systemic risk associated with large financial firms is well justified. From Long Term Capital Management to Lehman Brothers, shocks to individual firms have driven outsized aggregate losses. Additionally, financial firms exhibit unique features from other firms, where the behavior of market participants might exacerbate shocks contributing to fast-moving crises. However, given increasingly interconnected firms in diverse sectors, systemic risk might become increasingly present in the broader economy.

In fact, if one looks more closely at the 2008 crisis, the broader economy has already been a driver of systemic risk. In an effort to prevent a deeper crisis, Chrysler, Ford, and General Motors—the so-called “Big Three” American automakers—each received emergency loans to abate a larger crisis (Goolsbee and Krueger, 2015). Although the need for rescuing Ford and General Motors was apparent—both were under pressure from sharply decreased auto demand and had the potential to drive significant job losses and aggregate losses if they failed—the need

for rescuing Chrysler was, above all, about systemic risk. Leading up to the crisis, it had been estimated that of Chrysler's suppliers, 54 and 66 percent were also suppliers to Ford and General Motors, respectively (Goolsbee and Krueger, 2015). As a result, the risk of a Chrysler failure was the risk that it could pull down Ford, General Motors, or both, by first toppling shared suppliers. Seeing the significance of their interdependent supply chains, Ford publicly argued for the bailouts of its competitors, Chrysler and General Motors (Goolsbee and Krueger, 2015). The 2008 crisis—in which the financial sector emerged as the focal point for discussions on systemic risk—was, therefore, also a reminder that the risks of being too big or too interconnected to fail extend beyond a single sector.

The aggregate risk of micro-level shocks across the broader economy should come as no surprise. This precise risk has been studied in many ways. Notably, following the pioneering work of Leontief (1986), a common approach has been to measure the impact of micro-level shocks at the level of business sectors using input-output analysis and established sector-level tables of network flows. Although this approach enhances the understanding of aggregate risks, it lacks the resolution required for providing insight into individual firms. Understanding the contributions of individual firms to systemic risk would, instead, require a quantification of flows, inputs, and outputs at the firm level. This, however, does not yet exist.

To understand systemic risk in the broad economy and the contributions of individual firms, we introduce a novel approach that enables input-output analysis at the level of firms, using publicly available data to construct an observed interfirm production network. Although we gain insight through analysis of the observed network, we take the additional step of estimating a complete interfirm production network. This additional step employs statistical inference to expose likely unobserved supplier-customer connections in addition to a calibration procedure to estimate the potential flow of goods and services among firms in the U.S. economy. Finally, we consider the role of idiosyncratic firm-level shocks. Production networks provide a channel for economic contagion.

The main contribution of our work is methodological. Through data gathering, statistical inference, and network analysis, we advance the treatment of the microfoundations of aggregate shocks to measure the systemic risk posed by individual firms. In addition to the methodological contributions, our findings suggest that firms across diverse sectors might carry systemic risk. The contributions are described in the remainder of this report, which is organized as follows. In Chapter Two, we review the growing body of literature on systemic risk and production networks. In Chapter Three, we present a quantitative model of firm-level networks. In Chapter Four, we introduce data on observed supplier-customer linkages between firms and use basic network analysis to discuss the structure of the observed interfirm network. In Chapter Five, we present a methodology for estimating missing data on interfirm connections and thereby estimate a full production network for a subset of U.S. firms. In Chapter Six, we introduce idiosyncratic shocks at the level of firms to estimate contagion across firm networks and the potential systemic risk posed by individual firms. Finally, we discuss potential implications for policy and overall conclusions—including findings, limitations, and areas for future work—in Chapter Seven.



## Literature Review

---

### Systemic Risk and Contagion Is Not a New Idea

In this report, we contribute to the large body of research on systemic risk and contagion. The work on systemic risk builds on several notable contributions. Diamond and Dybvig (1983), for example, modeled bank runs with private information, excess risk, and moral hazard (i.e., the incentive toward excessively risky behavior because of an implicit or explicit promise of future protection). Others have focused on contagion, a metaphor to the disease-like spread of economic shocks, which Dornbusch, Park, and Claessens (2000) defined as “a significant increase in cross-market linkages” following an adverse shock. Kiyotaki and Moore (1995) presented a model in which cross-sector contagion follows from a vicious cycle of asset-price decline and declining investment. To understand contagion, many have explored the potential channels that allow shocks to spread. Building further on the Diamond and Dybvig (1983) model, Allen and Gale (2000) modeled liquidity preference, discussing optimal risk sharing and the role of interbank market structure. Notably, Allen and Gale (2000) brought attention to the role of network structure in facilitating the spread of shocks. Eisenberg and Noe (2001) studied credit contagion following firm defaults. Kodres and Pritsker (2002) further explored financial market contagion through a rational expectations model in which investors transmit shocks across markets through portfolio rebalancing. Furthermore, in Giesecke and Weber (2006), business partner networks serve as the channel for contagion and quantify the role of credit contagion on volatility and portfolio losses. As the literature on sys-

temic risk and contagion has grown to acknowledge the central importance of business networks, the need for estimating interfirm networks linkages has become more apparent.

## **Research Is Increasingly Exploring the Role of Networks in Financial Risk**

A growing segment of the contagion literature has focused on the role of networks. Acemoglu et al. (2012) built on the multisector real business cycle model of Long and Plosser (1983) to study the impact of sector-level input-output linkages and microeconomic shocks on macroeconomic fluctuations. A key contribution of their work stems from their finding that aggregate volatility depends on network structure and interconnectedness. Acemoglu et al. (2012) found that sector-level idiosyncratic shocks contribute to considerable aggregate volatility when there is significant sectoral asymmetry. Focusing on the financial system, Glasserman and Young (2015) studied the effect of financial system interconnectedness on systemic risk; Acemoglu, Ozdaglar and Tahbaz-Salehi (2015) studied the impact of network structure on the potential for contagion in the banking sector; and Bigio and La'O (2016) calibrated an input-output model at the sector level to study frictions across production networks. Importantly, the insights from detailed analysis of financial networks have underscored a notable gap in the literature: There is a need to understand firm-level network connections across the broad economy.

## **The Study of Risk Is Expanding to Include Production Networks**

More recently, the discussion of systemic risk and firm-level connections has begun to expand beyond financial networks. Carvalho (2014) argued that understanding production networks can elucidate changes in the macroeconomy. The 2011 Great East Japan Earthquake has provided an interesting example for the role of production networks.

MacKenzie, Barker, and Santos (2014) used a simulation of the 2011 Japanese earthquake in a model of severe supply-chain disruptions to estimate the impact on individual suppliers. Focusing on aggregate affects, Carvalho et al. (2016) used data from the Japanese earthquake to quantify the contribution of firm-level input-output linkages to the propagation of shocks downstream (forward linkages) and upstream (backward linkages). Within a general equilibrium model following the Long and Plosser (1983) and Acemoglu et al. (2012) setting, Carvalho et al. (2016) found that propagations played a significant role in aggregate losses. Using the same data set on customer-supplier relationships for Japanese firms, Arata (2018) included daily bankruptcy records to understand the propagation of bankruptcy across firm linkages, and found that while the majority of bankruptcies occur at isolated firms, firm-level network structures can play a significant role in macro shocks. These findings and others from the growing literature on production networks were summarized by Carvalho and Tahbaz-Salehi (2019).

Notably, there has been a lack of exploration into production networks at the level of firms within the United States, a gap explicitly highlighted in the summary of Carvalho and Tahbaz-Salehi (2019). This gap is largely because of a lack of firm-level data in the United States. As a result, Carvalho et al. (2016), Arata (2018), and others who focus on Japanese firm-level networks rely on firm-level linkages in Japan. Studies in the United States (e.g., Acemoglu et al., 2012) have primarily focused on sector-level networks. Barrot and Sauvagnat (2016) provided an exception and present a key suggestion for solving the challenge for U.S. data: They created a data set of firm linkages using a crucial U.S. regulation—Statement of Financial Accounting Standards (SFAS) No. 131—that requires publicly traded firms to report key customers and suppliers accounting for more than 10 percent of the firm’s total sales. Carvalho and Tahbaz-Salehi (2019) noted the challenge that these data present. The reporting requirements induce a double-selection bias by requiring reporting from only publicly traded firms and by observing connections typically between small and large firms. Yet the data has potential for providing substantial insight. According to Barrot and Sauvagnat (2016), the reported connections,

which are made public in a firm's Securities and Exchange Commission (SEC) 10-K filings, represent around 75 percent of total sales in Compustat, a popular database of financial statistics and information.

The data arising from SFAS No. 131 have become a key building block in exposing customer-supplier linkages, used both by others (e.g., Wu and Birge, 2014) and in the remainder of this report. In particular, we use this data with public reports to estimate a firm-level production network to investigate the micro origins of aggregate risks. The approach we take expands the utility of this data set by estimating firm-level networks while partially addressing the double-selection bias inherent to the data.



## The Model of Firm-Level Linkages

---

To begin the discussion of systemic risk at the level of firms, in this chapter, we briefly discuss the construction of a model of firm-level linkages that addresses an important gap in the aforementioned literature. We constructed a model similar to that of Acemoglu et al. (2012) that differs only in its focus on firms rather than sectors. The result is a model that formalizes the supplier-customer linkages between firms as input-output linkages, allowing for familiar analysis in the tradition of Leontief (1986). The complexity of this approach comes from the approximation of firm networks. We used this model to structure all subsequent analyses on estimating interfirm production networks and the systemic risk of individual firms.

To formalize a model of firm-level input-output linkages, we began by defining the economy as the tuple  $\mathcal{E} = \{F, \mathbf{W}, \epsilon\}$  where  $F$  is a set of firms,  $\mathbf{W}$  is a firm-level weighted-adjacency matrix (i.e., a network of firm-level input-output linkages), and  $\epsilon$  is a vector of idiosyncratic firm-level shocks. This construction is similar to that of Acemoglu et al. (2012) with the exception that in this construction, the economy is modeled at the level of firms rather than sectors.

Each firm in the economy produces output in the form of intermediate and final goods and services. We defined  $\mathbf{y} = [y_1 \dots y_n] \in \mathbb{R}^n$  as the firm output vector and  $\mathbf{d} = [d_1 \dots d_n] \in \mathbb{R}^n$  as the firm final demand vector. Furthermore, we defined  $\mathbf{W} = [w_{ij}]_{i,j} \in \mathbb{R}^{n \times n}$  as the firm-level weighted adjacency matrix, where  $w_{ij} \in \mathbb{R}^n$  are the flows from firm  $i$  to firm  $j$ . Firm-level output can thus be determined as the sum of flows and demand as follows:

$$\mathbf{y} = \mathbf{W}\mathbf{y} + \mathbf{D} \quad (3.1)$$

Then, using the standard approaches of input-output analysis, we further defined firm-level output as

$$\mathbf{y} = (\mathbf{I} - \mathbf{W})^{-1}\mathbf{d} = \mathbf{L}\mathbf{d} \quad (3.2)$$

where  $\mathbf{I}$  is an identity matrix and  $\mathbf{L} = (\mathbf{I} - \mathbf{W})^{-1}$  is the Leontief inverse matrix.

The true network of all supplier-customer linkages, however, is unknown. This true network of interfirm connections, as defined by adjacency matrix  $\mathbf{A} = \llbracket a_{ij} \rrbracket_{i,j} \in \{0,1\}^{n \times n}$ , where  $a_{ij} = 1$  if  $i$  is a supplier to  $j$ , and weighted adjacency matrix  $\mathbf{W} = \llbracket w_{ij} \rrbracket_{i,j}$  is only partially observed.<sup>1</sup> Because of the accounting standard, we observed only a partial network of direct connections  $\mathbf{D} = \llbracket d_{ij} \rrbracket_{i,j} \in \{0,1\}^{n \times n}$  where  $d_{ij} = 1$  if  $j$  is a known customer of  $i$  and  $d_{ij} = 0$  otherwise. That is, we observed a subset of all edges such that  $\mathbf{D} \times \mathbf{1} \leq \mathbf{A} \times \mathbf{1}$  where  $\mathbf{1}$  is an  $n$ -dimensional vector of 1s. Furthermore, uncertainty is defined by the probability  $p_{ij}$  of a directed edge from  $i$  to  $j$  such that  $p_{ij} = \Pr(a_{ij} = 1)$ .

Chapter Four describes the true, unknown, interfirm network as measured using available financial data and statistical inference. It describes the partial network of observed connections  $\mathbf{D}$ , and Chapter Five describes missing data on network connections to approximate the network described by  $\mathbf{A}$ . Finally, Chapter Six builds on the model presented here to present a propagation model that estimates contagion across firm networks.

---

<sup>1</sup> The adjacency matrix  $\mathbf{A}$  consists of binary connections between firms and thereby makes up a network of directed edges, and the weighted adjacency matrix  $\mathbf{W}$  consists of directed and weighted edges and thereby represents the full interfirm network with the flow of goods between firms.

## **Observed Interfirm Production Network**

---

In this chapter, we present an approach that builds on the work of Barrot and Sauvagnat (2016) and uses SFAS No. 131 reporting requirement to produce an expanded data set of observed interfirm connections. Additionally, we use this chapter to discuss a descriptive analysis of the observed network and its implications. Using conventional methods from network analysis, we identified areas of dense connections between firms and firms with high levels of network centrality (i.e., firms of significant network importance). This chapter includes a discussion of specific measures of network centrality, their challenges applicable to production networks, and the need for expanded analysis that considers the economically weighted significance of each firm in the interfirm network.

### **Data Sources and Techniques**

As previously noted, SEC regulation SFAS No. 131 requires public firms to report information about business segments and customers that represent more than 10 percent of consolidated sales or revenues. We constructed a data set of firm-to-firm linkages by sweeping the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system for each firm filing with the SEC. Pulling data from public records from Standard and Poor's (S&P's) Capital IQ database, we obtained the names of each firm's significant customers and associated sales from the first quarter of 2015 to the first quarter of 2019. We used the SEC Financial Statements and Note data sets, exploiting the

reported “Concentration Risk” segments to identify these significant supply partners. We then searched this data set for firms that note that none of their business relationships constitute more than 10 percent of net sales or revenue, and marked these relationships as representing insignificant supply flows.

The data set of observed connections yields more firm linkages than might be expected. First, although some companies might report few or even zero significant customers or suppliers (e.g., large firms with no segments exceeding 10 percent of sales), these companies’ connections to other nonsignificant firms might be reported if the linkages account for more than 10 percent of the nonsignificant firms’ sales. For example, even if Apple does not indicate any significant customers or suppliers on its financial reports, many of its customers and suppliers might report Apple on their financial reports, resulting in a network of observed connections that includes Apple. Second, some firms go beyond required reporting, revealing suppliers and customers below the 10-percent threshold. Third, although private companies are not required to report linkages, public companies might be required to report linkages to private firms. Importantly, taken together, we cannot determine which firm linkages have been systematically omitted from the observed network.

Although some previous studies (e.g., Barrot and Sauvagnat, 2016) have addressed this missing data problem by restricting their samples to public firms, we use an intermediary step. We wrote a custom algorithm to search the SEC’s EDGAR engine to find the XML (eXtensible Markup Language) address of each firm’s annual filing, query the address, and extract Exhibit 21 of the document. We then used the contents to construct a database of parents to subsidiaries. By assigning private companies to their public parent whenever possible, we elevated our analysis to the parent level, both augmenting our data set and helping us ignore intrafirm supply relationships, which we believe pose a less significant risk to the broader economy.

The remainder of this chapter summarizes the observed data on firm-level networks from financial reports. Although these networks are partial, we analyzed their structure to provide insight into the

potential for contagion and systemic risk from firms across all sectors of the economy.

## Description of Observed Network

From the aggregation of public data, we found 124,786 connections across 61,156 firms. Although this represents a large set of connections, we restricted this data set to supplier-customer relationships among firms for which we have both total revenue (i.e., output) and total costs of goods and services sold (COGSS)<sup>1</sup> to estimate the strength of input and output linkages. Unfortunately, this data is not available for the majority of private firms. As a result, our final observed network connection was restricted to 20,477 connections across 5,821 firms.

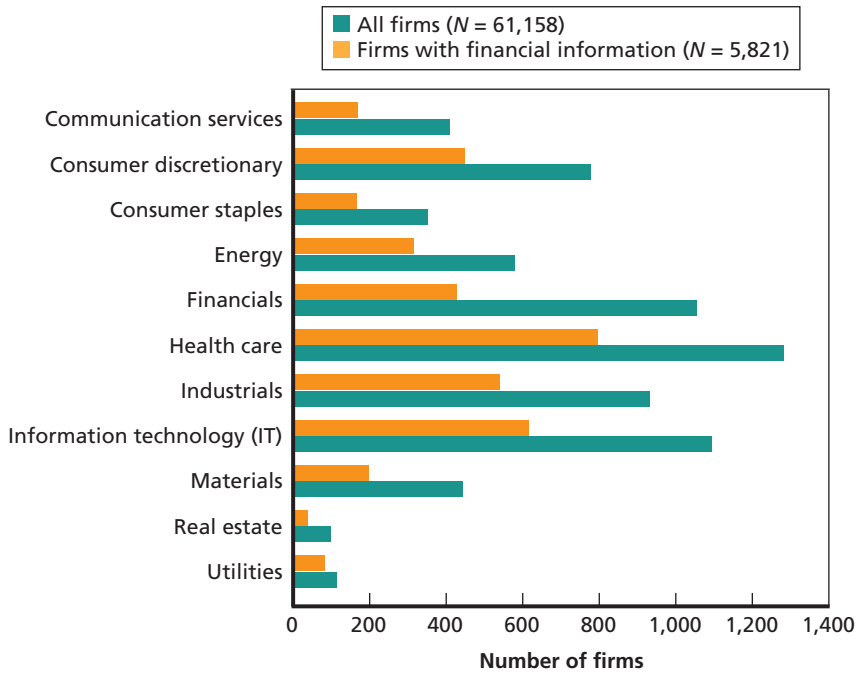
Figure 4.1 shows that the distribution of our obtained subsample by sector looks similar to the overall distribution. Indeed, when we excluded the “Financials” sector, the  $p$ -value for the Chi-Squared Test of Homogeneity is 0.754.<sup>2</sup> When we included the Financials sector, this value falls to 0. We suspect that we captured relatively fewer financial firms in our subsample because the selection criteria are: (1) to be identified as customer or supplier, and (2) have COGSS or revenue reported to the SEC in a way that we observe (via EDGAR or S&P). In the financial sector, in particular, it is very often the case that firms we observe do not report to the SEC and neither do their parent firms. This is because a large number are: (1) foreign, (2) have foreign parents, or (3) are banks that report to their regulator rather than to the SEC (our data source). Future research could supplement our insight into this sector by exploring how to extract additional information from the Federal Deposit Insurance Corporation database.

---

<sup>1</sup> The COGSS includes all variable costs associated with production, such as labor, materials, and shipping, but does not include costs that are directly tied to production, commonly called *overhead costs*.

<sup>2</sup> This test was performed using the “chisq.test” function of the “stats” package in R. For a detailed discussion, see Ricci (2005).

**Figure 4.1**  
**Distribution of Data Coverage Across Sectors**

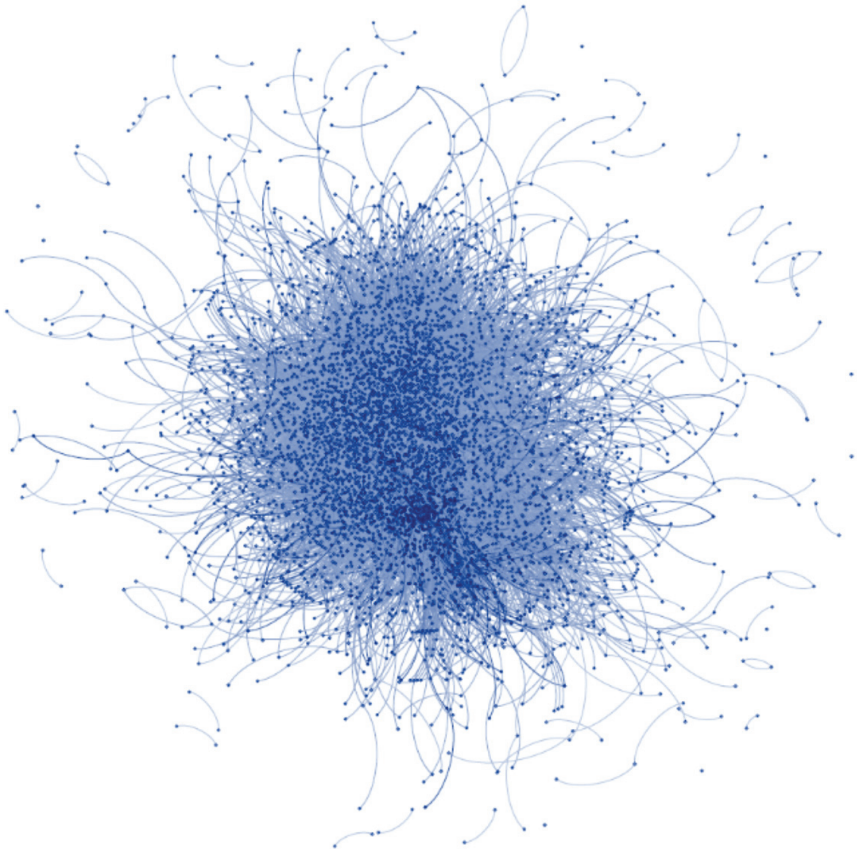


NOTE: This figure shows the relative number of firms with the required financial information (revenue and COGSS) included in the 5,821-firm data set versus all observed firms (with and without financial information) across Global Industry Classification Standard business sectors.

The supplier-customer linkages extracted from the SEC EDGAR database produce a sparse network of interconnected firms. Figure 4.2 displays the resulting network graph. Each node represents an individual parent firm while each directed edge represents an observed supplier-customer linkage.

The network graph in Figure 4.2 reveals several notable features of interfirm connections. The network structure itself is quite sparse; many firms lie on the periphery, with very few observed connections. As firms move toward the center, their connections become quite dense. Nonetheless, the absolute number of firms displayed in

**Figure 4.2**  
**Observed Production Network Graph**



NOTE: Visualization of network graph with 20,477 observed directed edges across 5,821 firms.

Figure 4.2 requires more-detailed analysis to fully characterize the structure of the observed network.

### Analysis of Observed Network Structure

The observed network  $\mathbf{D} = \{d_{ij}\} \in \{0,1\}^{n \times n}$  (recall that  $d_{ij} = 1$  if  $j$  is a known customer of  $i$  and  $d_{ij} = 0$  otherwise) can be summarized by several network statistics. First, consider the degree of each node (i.e., parent firm)  $i$  in the network. The degree—the sum of connections into and out of node  $i$ —is calculated as follows:

$$\delta_i = \delta_i^{IN} + \delta_i^{OUT} = \sum_j d_{ji} + \sum_j d_{ij} \tag{4.1}$$

where the sum of edges into node  $i$ ,  $\delta_i^{IN}$ , is its indegree and the sum of all edges out of node  $i$ ,  $\delta_i^{OUT}$ , is its outdegree. In the context of economic contagion, the degree of node  $i$  indicates the number of firms that would be affected immediately following a shock to node  $i$ .

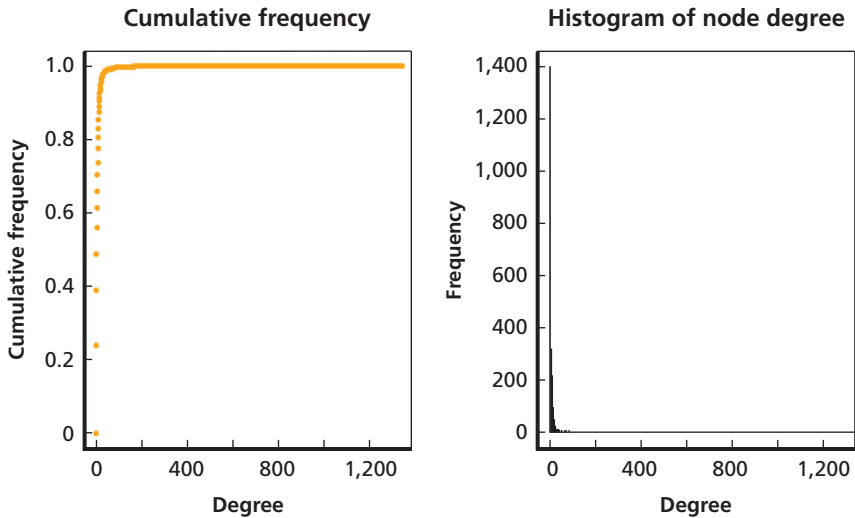
Observed node degrees are summarized in Table 4.1. The average degree in the network  $\mathbf{D}$  is 7 with a median of 3. That is, on average, each firm is connected to seven other firms. The maximum observed node degree is 1,334, while the minimum is 0, reflecting unobserved connections. Removing nodes without connections, the average degree of connected nodes is 9.3, with a median of 5. As the difference in mean and median reflect, the distribution of node degree is quite skewed. Figure 4.3, which shows this distribution by cumulative frequency and

**Table 4.1**  
**Summary of Node Degree Statistics**

Measure	Value
Average degree, $\bar{\delta}_i$	7
Median degree, $\tilde{\delta}_i$	3
Maximum degree, $\max\{\delta_i\}$	1,334
Minimum degree, $\min\{\delta_i\}$	0



**Figure 4.3**  
**Node Degree Distribution**

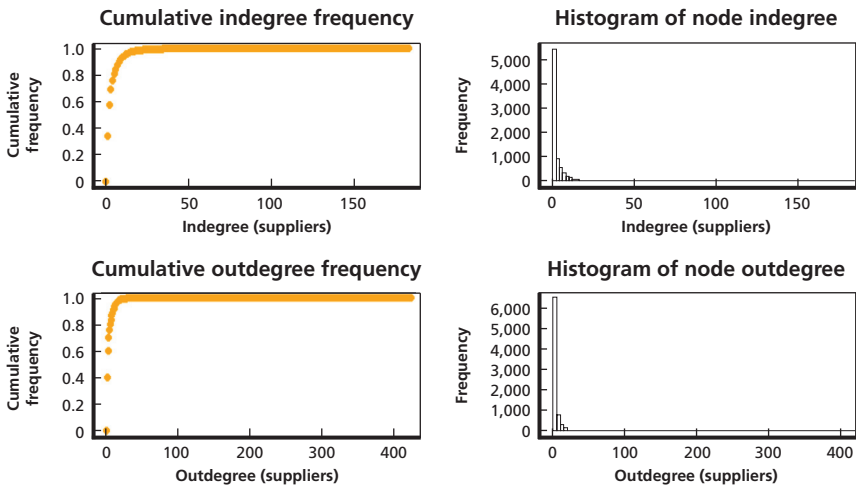


a histogram of node, demonstrates that a few firms represent the vast majority of observed firm connections.

The distribution of degree varies further by direction. Figure 4.4 shows distributions of both indegree (total known suppliers) and outdegree (total known customers). Both exhibit similar skewed distributions as total degree; however, outdegree is more skewed than indegree. This is partially reflected in their maxima—the maximum observed number of suppliers to one firm (indegree) is 224, whereas the maximum number of customers (outdegree) is 1,282—but also in their centers of mass. The implication of the observed network characteristics is a network with more known customers than suppliers. This fact is intuitive for production networks in which the relative level of indegree and outdegree is likely to depend on firms and their sectors.

In fact, firms with high indegree are predominantly classified as retail trade; wholesale trade and transportation; communications; and electric, gas, and sanitary service sectors, with the highest indegrees from the services and manufacturing sectors. These are big compa-

**Figure 4.4**  
**Node Indegree and Outdegree Distribution**



nies (e.g., Amazon, Walmart, AT&T, Apple, Alphabet [Google], Ford, Chevron, Comcast, Exxon Mobile, Cardinal Health) that have some firm-level customers but mostly sell to households, a type of customer unobserved within production networks, by definition. That is, outdegree only counts firm-level customers, not households. Furthermore, firms with 0 outdegree and relatively high indegree are predominately represented in the retail trade sectors, which is unsurprising for firms whose suppliers are predominately other firms and whose customers are predominately households. The relatively high levels of indegree suggest that firms in these groups have a disproportionately high upstream impact following adverse shocks but might have considerable welfare effects because of final demand disruptions, though we are unable to estimate the effects because we cannot separate intermediate from final demand in the data.

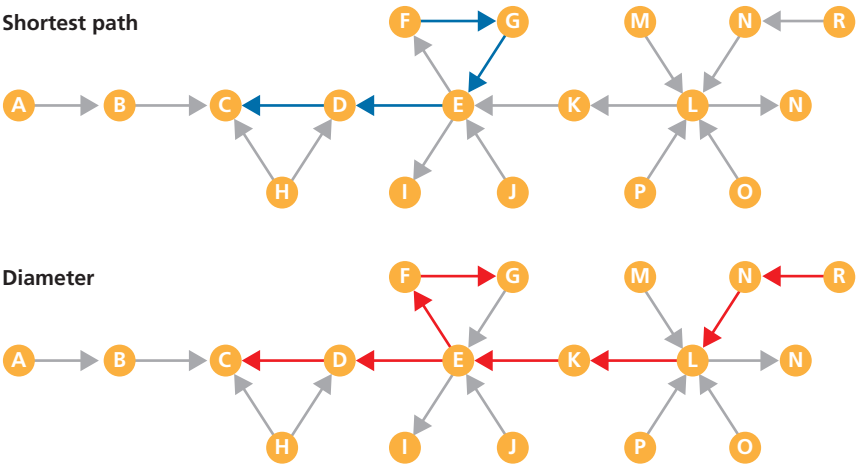
Firms with high average outdegree predominately operate in finance, insurance, and real estate; services; and transportation, communications, electric, gas, and sanitary service divisions. In fact, the top ten companies with the highest outdegree score are all financial institutions. Despite the fact that this sector represents only 11.5 per-

cent of all nodes, it accounts for 33 percent of all edges. Providers of communications and IT solutions (e.g., Microsoft, AT&T, Alphabet, Dell, Cisco, IBM) also represent high levels of outdegree. Furthermore, firms with zero indegree and fairly high outdegree include not only financial institutions, but also manufacturers and providers of IT solutions mostly for other businesses (supply chain management software, commerce, and artificial intelligence software). The relatively high values of outdegree suggest that these firms have a disproportionately high downstream impact following adverse shocks.

Going beyond degree, the distances between nodes shed further light on the structure of the production network. The *distance* between node  $i$  and node  $j$  (length of a path) is defined as the number of steps (connections) between these two nodes. Then, the length of the shortest directed path between two nodes is defined as the *geodesic distance* between two nodes. Furthermore, the longest geodesic distance in the network is defined as the network's *diameter*. That is, the diameter of a network is its longest shortest path. In reality, the network's diameter provides an indicator of how firms are dispersed across the economy by measuring the economy's longest (observed) supply chain. These distances are visualized in the simplified graph in Figure 4.5. If a node degree shows how wide the impact of a shock would be (first order influence), then the distance shows how far it could potentially propagate.

In any given directed network, the total number of possible connections (excluding loops) is given by  $N(N-1)$ , where  $N$  is the number of nodes. Given that the number of nodes in network **D** is 5,821, the total number of possible connections is 33,878,220. However, almost 70 percent (23,513,697) of possible connections are either nonexistent or unobserved, and 0.06 percent of possible connections are direct links (i.e., they have the geodesic distance of 1). The corresponding distribution of geodesic distances between nodes in network **D** is presented in Figure 4.6, which reflects a mean geodesic distance of 4.8 and maximum distance (i.e., network diameter) of 14, which can be interpreted as the length of the average supply chain and the longest observed supply chain in the network, respectively. Interestingly, Arata (2018) estimates an average geodesic distance of 3.9 for the Japanese

**Figure 4.5**  
**Network Distance, Shortest Path, and Diameter**

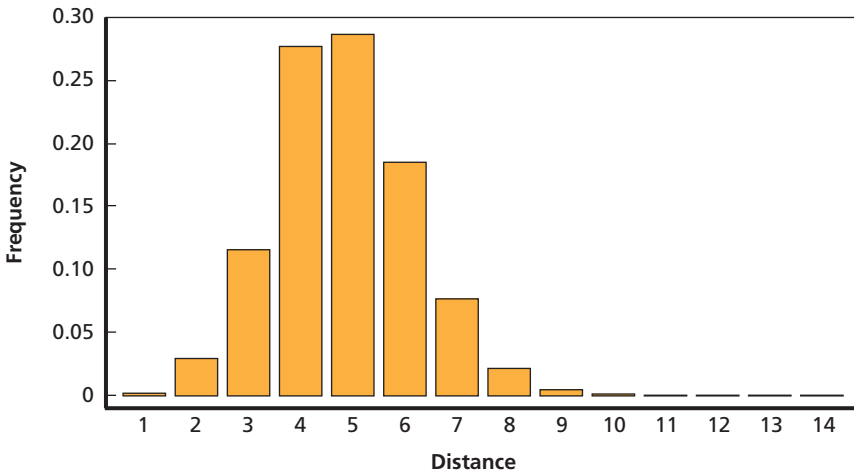


NOTE: The shortest path between nodes F and C (path F-G-E-D-C) means geodesic distance between F and C is 4. The longest geodesic distance (paths from R to C and R to G) is 6, the diameter of this network.

economy. Although this observation suggests that Japanese firms are, on average, closer to other firms than in the United States, this is hard to say exactly. Knowing that our data set of U.S. firms is only partial, it is very likely that increasing our known observations could result in shorter distances. Further empirical studies of U.S. interfirm networks could explore potential differences in network composition by country.

However, although the structure of the network (degree and distance) improves our understanding of how wide and far a given shock might propagate, the fundamental questions of systemic risk are whether some nodes are so structurally important that they could cause disproportionate losses from firm-level shocks. Traditional network analysis would approach this question by ranking nodes in order of importance as determined by multiple centrality indices (for example, see Bonacich [1987]). Depending on how the importance of a given node is defined, different centrality measures will apply. The most commonly used centrality measures include the following:

**Figure 4.6**  
**Distribution of Geodesic Distances in Network D**



NOTE: This figure shows the distribution of firms according to their geodesic distances. The y-axis shows the frequency, or percentage of firms in the sample corresponding to each geodesic distance shown through the x-axis.

1. *Degree centrality* indicates which node directly influences the most other nodes: For directed networks, this is typically defined separately for incoming and outgoing connections; i.e., indegree and outdegree. As a result, degree centrality identifies nodes with the biggest number of direct links to other nodes.
2. *Eigenvector centrality* indicates which node directly and indirectly influences the biggest number of other nodes; high eigenvector centrality is given to nodes that are connected to many other nodes,<sup>3</sup> which are also connected to many other nodes, and so on. Eigenvector centrality would be a good centrality measure (if a shock propagated indefinitely across a network) without attenuation (a similar centrality measure that takes potential attenuation into account is the Katz centrality).

<sup>3</sup> Eigenvalue centrality  $\mathbf{x}$  is given by  $A^T \mathbf{x} = \lambda \mathbf{x}$ , where  $\lambda$  is a diagonal matrix of eigenvalues (Bonacich and Lloyd, 2001).

3. *Closeness centrality* indicates which node has the shortest distance to all other nodes in a network and, consequently, if hit by a shock, would spread it faster than others.
4. *Betweenness centrality* indicates which node lies on the largest number of shortest paths between other nodes. This measure is usually used to identify nodes that could serve as bridges between communities. Betweenness disregards nodes that are at the beginning or at the end of a path.

However, the case of production networks has notable features that challenge the direct application of each measure. For example, the shocks in the economy will not spread infinitely (e.g., Arata [2018]) but instead attenuate while moving across the supply chain (e.g., Carvalho et al. [2016]). We, therefore, considered a separate approach in Chapter Six that builds on the input-output model presented in Chapter Three to measure economic centrality in production networks.

## Estimating Production Networks Through Inference

---

Although the observed network linkages between suppliers and customers provide insight into production network structures and begin to capture the interconnectedness of U.S. firms, the data are only a partial network. As previously mentioned, observations of network connections occur only if they meet reporting requirements for public firms, leading to a double selection bias that Carvalho and Tahbaz-Salehi (2019) note as being the primary challenge for firm-level analysis. The result gives rise to two distinct problems: (1) inferring observed supplier-customer linkages and (2) estimating the economic strength of those linkages, both observed and inferred. In this chapter, we introduce a methodology for addressing these questions as a two-step process. Although the analytic methods presented in this chapter represent an advancement in the treatment of interfirm production networks, they should be viewed as a first step.

### Inference for a Complete Network

We estimated the potential network connections as probabilities using a gravity-type model using known sector-to-sector relationships within the data. One formulation of the problem would model the missing edges as random missing information in an observation of the network. This suggests the use of a statistical imputation method to infer estimates for the unobserved edges.

### **Simplifying Assumption: Unobserved Firm Connections Are Missing at Random**

Data on observed firm connections are generated by firms reporting connections of significant suppliers and customers and thereby appear filtered. However, we might observe a connection because it exceeded the reporting threshold of the customer, the supplier, or both. Consequently, the true filtering process is currently unknown. We therefore treat the filtering process as random, leading to the fundamental assumption that unobserved connections are missing at random.

Although we know that this is not the case, we were unable to identify systematically which firm linkages were left out of the observed network. Using the observations of Barrot and Sauvagnat (2016), we might observe more small-to-large firm linkages than any other type, but how this affects the analysis is unknown.

We used a logistic regression with the full network of possible connections to identify a probabilistic network that can be used to construct potential networks for analysis. This approach relied on the assumption that all unobserved firm connections are randomly distributed within the true network  $\mathcal{A} = \llbracket a_{ij} \rrbracket_{i,j}$ . Because large firms are likely to have many customers and suppliers, they might be less likely to report relationships with other large firms using SFAS No. 131. Because small firms are likely to have a small number of relationships, these small firms can be used to identify the potential network of larger firms using the accounting standard.

In addition to the identified relationships between firms, we used the revenue of the customers and suppliers, the product of the revenues, and the 1-digit Standard Industry Classification (SIC) code. Using these 1-digit codes, we constructed the full factorial interaction between every customer sector with every supplier sector. This allowed us to identify the relative importance of sectoral relationships in identifying individual firm relationships. The logistic regression is characterized by:

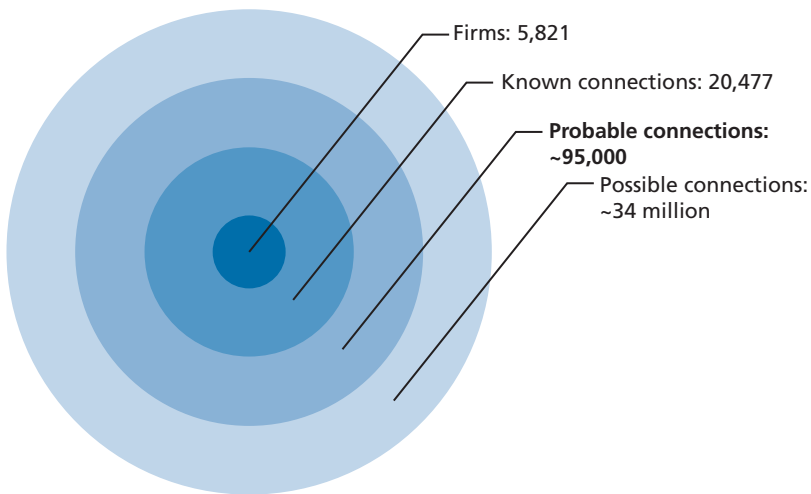
$$p_{ij} = \Pr(a_{ij} = 1) = \frac{1}{1 + e^{-(\mathbf{x}'\beta)}} \quad (5.1)$$



where  $\mathbf{X}$  is a matrix of revenues and sectoral relationships across observations and  $\beta$  is a vector of coefficients; recall that  $p_{ij} = \Pr(a_{ij} = 1)$  is the probability of a connection between  $i$  and  $j$ . Specifically, given the likelihood that a given relationship is influenced by the size of the customers, suppliers, and their respective sectors, we estimate Equation 5.1 by regressing the observed network on customer revenue, supplier revenue, and a set of binary dummy variables on customer and supplier SIC sector codes, the sector codes used by our underlying data.<sup>1</sup>

The results of this estimation are illustrated in Figure 5.1. Using approximately 20,000 known connections across nearly 6,000 firms, we estimated that there were potentially 95,000 connections not

**Figure 5.1**  
**Probable Connections Uncovered by Logistic Regression**



<sup>1</sup> Although our underlying data source is SIC-based, the North American Industry Classification System (NAICS) is the current standard for industry classification and was developed by the Office of Management and Budget to replace SIC. For conversion, NAICS publishes a crosswalk between the two sector codes.

observed in the data.<sup>2</sup> That is, there are fivefold more probable connections than we directly observed from data.<sup>3</sup>

## Network Calibration

In this chapter, we introduce an approach that calibrates a possible network structure to data on the real economy. To do so, we considered the total inputs (COGSS)  $c_i$  and outputs (sales)  $y_i$  of each firm and an inferred adjacency matrix,

$$\hat{A} = \left[ \left[ \hat{a}_{ij} \right] \right]_{i,j}$$

given probabilities  $p_{ij}$ , to estimate a full weighted adjacency matrix

$$\hat{W} = \left[ \left[ \hat{w}_{ij} \right] \right]_{i,j}.$$

Furthermore, we defined residual output  $y_{ir}$  as the amount of firm output  $y_i$  that is not used as inputs to other known firms and is either sold to unknown firms or is final demand. The resulting calibration indicated the most efficient allocation of inputs and outputs across likely networks of supplier-customer connections. We defined the calibration procedure as an optimization model below.

First, the flow of output from each firm to others in its network is constrained by total output (sales). That is, the output of each firm  $i$  must either be sold to other known firms  $j$  or to  $y_{ir}$ . This relationship can be described as follows

$$y_i = \sum_{j \in F} w_{ij} y_j + y_{ir} \quad \forall i \in F \quad (5.2)$$

---

<sup>2</sup> We defined a potential connection as a connection that has greater than a 50-percent probability of occurring using the regression results.

<sup>3</sup> The regression results, estimated probabilities, and underlying data are available from the authors on request.

where  $w_{ij}y_i = y_{ij}$  is the flow of outputs from  $i$  to all other connected firms  $j$ , and  $y_{ir}$  is the flow of outputs to residual. That is, all of firm  $i$ 's output is consumed. Similarly, the inputs (COGSS) for every firm  $i$ ,  $c_i$ , are equal to the sum of its inputs from each firm  $j$ , the product  $w_{ji}y_j$ , and its inputs from residual  $y_{ri}$  as follows:

$$c_i = \sum_{j \in F} w_{ji}y_j + y_{ri} \quad \forall i \in F \quad (5.3)$$

That is, all firms' inputs must come from somewhere. Importantly,  $y_{ri}$  and  $y_{ir}$  should be positive for most. Additionally, if no connection between two firms  $i$  and  $j$  was observed such that  $a_{ij} = 0$ , then it must be that  $w_{ij} = 0$ . This constraint can be written as follows:

$$w_{ij} \leq a_{ij} \quad \forall i, j \in F \quad (5.4)$$

Similarly, if a connection between two firms  $i$  and  $j$  was observed such that  $a_{ij} = 1$ , then it must be that  $w_{ij} > 0$ . The condition that  $a_{ij} = 1 \Rightarrow w_{ij} > 0$  is equivalent to the following constraint:

$$w_{ij} \geq \varepsilon a_{ij} \quad \forall i, j \in F \quad (5.5)$$

where  $\varepsilon$  is a small positive value near zero. Furthermore, the sum of firm  $i$ 's output weights into all firms  $j$  and cannot exceed 1. That is, firms cannot sell more than they produce:

$$\sum_{j \in F} w_{ij} \leq 1 \quad \forall i \in F. \quad (5.6)$$

Finally, the objective of this calibration exercise was to find an efficient allocation of output across observed network connections  $a_{ij}$ . The most efficient is the set of allocations that minimize the sum of all residuals as follows:

**Table 5.1**  
**Results of Optimization Model Runs**

Result Description	100 Firms	500 Firms	1,000 Firms
Solve status	Optimal	Optimal	Optimal
Memory allocated	19.69 Mb	19.69 Mb	2,133.23 Mb
Run time	0:00:00.069	0:00:06.138	3:56:17.646

NOTE: Mb = megabytes. This table shows the results of calibrating the model by optimization with 100, 500, and 1,000 firms included. Although it has not been rigorously studied, the results appear to convey computational complexity in logarithmic time.

$$\min_{w_{ij}} \sum_{i \in F} y_{ir} + y_{ri}. \tag{5.7}$$

That is, the calibration exercise results in a linear program where Equation 5.7 is minimized subject to constraints in Equations 5.2 through 5.6.

The results of this calibration exercise with actual firm data and estimated networks are summarized in Table 5.1. We tested the results of the optimization including three partitions of the data: the top 100, 500, and 1,000 firms by revenue in our sample. The results reflect a complex computational problem that, although fast at the level of 100 firms, became computationally expensive at the level of 1,000 firms. We used the results of the 1,000-firm test run in subsequent analysis on the systemic risk of individual firms. Although restriction narrowed the sample, it offered improved tractability for analysis over the full sample. The computational intensity together with time constraints meant that we were able to estimate only a subnetwork of the estimated network. Future work will focus on expanding the ability to estimate larger networks through better computational methods that reduce the solve time while allowing for multiple draws from the probable network.

## Contagion Across Firm Networks

---

In this chapter, we return to the original model developed in Chapter Three and consider the potential for contagion across firm supply chain networks. We started by introducing a vector of idiosyncratic firm shocks  $\epsilon = [\epsilon_1 \dots \epsilon_n] \in \mathbb{R}^n$ . In this vector, each element  $\epsilon_i$  represents an output shock isolated to an individual firm  $i$  where, following a shock to firm  $i$ , the change in output is the product  $\epsilon_i y_i$ . Furthermore, given the  $\epsilon_i$  shock to individual firm  $i$ , the flow of inputs to all depending firms  $j$ , defined as  $y_{ij}$ , are reduced by

$$\epsilon_i \left( \frac{y_{ij}}{\sum_f y_{if}} \right),$$

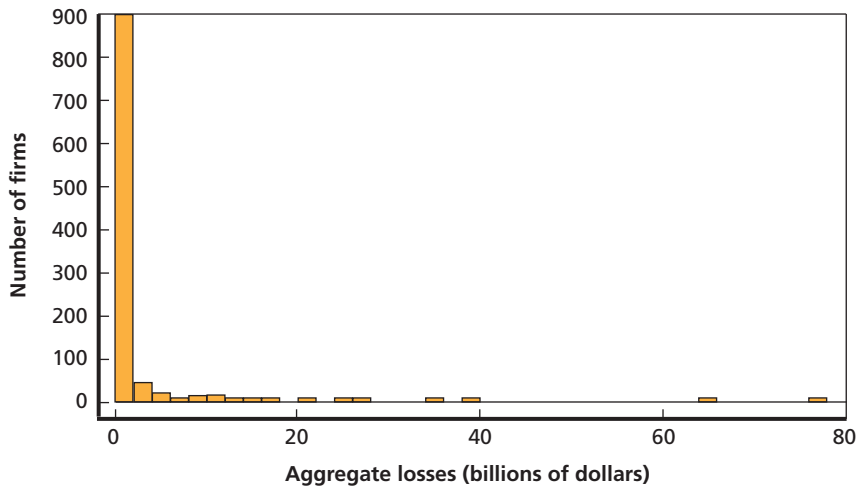
where  $\sum_f y_{if}$  is the sum of inputs used by all firms  $f$  from  $i$ . Let  $\mathbf{Z}$  be an  $n \times n$  matrix where  $z_{ij}$  is the proportion of firm  $i$ 's input that comes from firm  $j$ . Furthermore, recall that  $\mathbf{y}$  is a  $n \times 1$  vector where the  $i$ 'th element  $y_i$  is firm  $i$ 's output and  $\epsilon$  is a  $1 \times n$  vector where the  $i$ 'th element  $\epsilon_i$  is the proportion of firm  $i$ 's revenue lost because of a shock. Therefore, the total loss is given by:

$$\Delta y = \epsilon \left( \sum_{j=1}^{\infty} z_{ij} \right) R \quad (6.1)$$

To better understand Equation 6.1, suppose that the shock did not propagate beyond the initially shocked firm. Then, the term  $\epsilon Z y$  represents the total economic loss. Specifically, the term  $Z y$  is a vector that gives each firm's output; thus, multiplying that vector by  $\epsilon$  gives the total economic losses if the shock only propagated one step. Now, suppose the shock propagates two steps. Recall that  $\epsilon$  is a vector that gives the proportion of output lost because of an initial idiosyncratic shock. The term  $\epsilon Z$  gives the proportion of output lost to all firms because of the one step propagation of the initial shock  $\epsilon$ . Therefore, all second-step losses are given by  $\epsilon Z Z y$ , and the total first and second step losses are given by  $\epsilon Z y + \epsilon Z Z y$ . Because we assume the propagation continues indefinitely, the total losses are given by  $\epsilon Z y + \epsilon Z Z y + \epsilon Z Z Z y \dots$ , which, when written in summation notation, is Equation 6.1.

Figure 6.1 shows the distribution of total losses for a 1-percent shock to an individual firm's output (revenue), for all firms.<sup>1</sup> That is,

**Figure 6.1**  
**Distribution of Aggregate Losses from Idiosyncratic Shocks**



NOTE: This figure shows the distribution of total economic losses following a 1-percent output shock to individual firms where the y-axis displays the number of firms and the x-axis the value of each shock in billions of dollars.

<sup>1</sup> This is equivalent of taking the firm offline for 1 percent of the year.

we shock each firm by decreasing output by 1 percent, estimating the total losses to the network one firm at a time. The consideration of the 1 percent is large but not unreasonable; a natural disaster or extreme weather event provides an illustrative example. Importantly, the qualitative nature of the histogram does not depend on the size of the initial shock. Instead, different initial shocks would just scale the horizontal axis. As the plot shows, the distribution is skewed toward zero with a long tail. Specifically, the median total loss is approximately \$50 million if the maximum total loss is greater \$70 billion, which is more than 100 times the median. This illustrates the fact that many firms are relatively unimportant when considering systemic outages, but there are a small number of firms of critical importance.

Table 6.1 displays an estimation of the top 20 firms in terms of the total economic loss that a 1-percent shock to their revenue would cause. The dollar value of losses resulting from each shock represents a single point estimate of the direct and indirect losses resulting from propagations along the estimated network structure on which the firm relies. Although this value is not intended to have predictive power, it is reflective of the potential for economic contagion following a shock to a single firm. As with the results shown in Figure 6.1, this table demonstrates the fact that some firms might hold significantly higher systemic risk than others: Table 6.1 illustrates a steep drop-off as the top two firms are almost twice as central as firms five through 20. Furthermore, Table 6.1 conveys considerable heterogeneity; the majority of the most-central firms lie outside the financial sector (the traditional focus of systemic risk conversations), and they vary across diverse business sectors. Additionally, Table 6.1 shows that the potential for contagion is driven not only by revenue; instead, firms have disproportionate impacts on the overall economy following network effects.

In comparison, Table 6.2 shows the top 20 firms according to aggregate losses per unit of their revenue. Although many of these firms are relatively small (note the change in units from billions of dollars to millions), their loss-to-revenue ratios are considerably higher than those of Table 6.1. Although a shock to Amazon has an effective aggregate multiplier of 0.54, a shock to GoDaddy has a multiplier of 18. The difference is not entirely surprising; despite its revenue,

**Table 6.1**  
**Top 20 Firms with the Highest Estimated Total Network Losses**

Firms	Losses (\$, Billions)	Revenue (\$, Billions)	Ratio of Losses to Revenue	SIC Sector
Amazon.com, Inc.	77	142	0.54	Miscellaneous Retail
Comcast Corporation	65	85	0.76	Communications
UnitedHealth Group Inc.	40	155	0.26	Insurance Carriers
Walmart, Inc.	34	500	0.07	General Merchandise Stores
Prudential Financial, Inc.	28	45	0.62	Insurance Carriers
Apple, Inc.	27	160	0.17	Electronic and Other Electrical Equipment and Components
CVS Health Corporation	27	131	0.21	Miscellaneous Retail
AT&T, Inc.	26	85	0.30	Communications
The Allstate Corporation	26	38	0.68	Insurance Carriers
Amerisourcebergen Corporation	25	155	0.16	Wholesale Trade— Nondurable Goods
Bank of America Corporation	23	95	0.24	Depository Institutions
Exxon Mobile Corporation	22	119	0.18	Petroleum Refining and Related Industries
Citigroup, Inc.	21	93	0.23	Depository Institutions
International Business Machines Corporation	21	79	0.26	Business Services
The Progressive Corporation	20	27	0.73	Insurance Carriers
Anthem, Inc.	19	90	0.21	Insurance Carriers
McKesson Corporation	18	78	0.23	Wholesale Trade— Nondurable Goods
Tech Data Corporation	17	31	0.55	Wholesale Trade— Durable Goods
General Dynamics Corporation	17	31	0.54	Transportation Equipment
The Boeing Company	17	16	1.03	Transportation Equipment



**Table 6.2**  
**Top 20 Firms with Highest Estimated Losses Per Unit Revenue**

Firms	Losses (\$, Millions)	Revenue (\$, Millions)	Ratio of Losses to Revenue	SIC Sector
GoDaddy, Inc.	4.02	0.22	18.00	Business Services
PRA Health Sciences, Inc.	1.95	0.23	8.63	Engineering, Accounting, Research, and Management Services
Charles River Laboratories International, Inc.	1.60	0.19	8.63	Engineering, Accounting, Research, and Management Services
Moody's Corporation	2.00	0.42	4.76	Business Services
C&J Energy Services, Inc.	1.48	0.33	4.52	Oil and Gas Extraction
Sprague Resources LP	1.13	0.29	3.97	Wholesale Trade—Nondurable Goods
Hologic, Inc.	1.24	0.31	3.97	Measuring, Photographic, Medical, and Optical Goods, and Clocks
Catalent, Inc.	8,458.04	2,299.70	3.68	Chemicals and Allied Products
Cleveland-Cliffs Inc.	2.09	0.70	2.98	Metal Mining
Newfield Exploration Company	1.53	0.53	2.88	Oil and Gas Extraction
The PNC Financial Services Group, Inc.	4,634.00	1,887.94	2.45	Depository Institutions
Superior Energy Services, Inc.	1.76	0.75	2.34	Oil and Gas Extraction
Targa Resources Corporation	10,332.12	4,409.65	2.34	Electric, Gas, and Sanitary Services
EOG Resources, Inc.	12,246.57	5,615.15	2.18	Oil and Gas Extraction
Workday, Inc.	3,373.58	1,744.78	1.93	Business Services
Diebold Nixdorf, Inc.	737.14	384.94	1.91	Industrial and Commercial Machinery and Computer Equipment

**Table 6.2—Continued**

Firms	Losses (\$, Millions)	Revenue (\$, Millions)	Ratio of Losses to Revenue	SIC Sector
AMC Entertainment Holdings, Inc.	4,684.56	2,572.92	1.82	Motion Pictures
Wabash National Corporation	2,808.50	1,767.20	1.59	Transportation Equipment
Verisk Analytics, Inc.	1,569.65	1,074.53	1.46	Insurance Agents, Brokers and Service
Portland General Electric Company	0.86	0.60	1.43	Electric, Gas, and Sanitary Services

NOTE: This table shows the top 20 firms according to per revenue aggregate losses following a 1-percent idiosyncratic shock. Firm losses and revenue (shown in millions of dollars) are displayed to the left of their ratio and the firms’ SIC business sector.

GoDaddy is a large internet domain registrar that provides an essential service keeping the digital storefronts of numerous companies active. As a result, a shock to GoDaddy (which could follow from natural disasters or cyberattack) might have a disproportionate effect. As in Table 6.1, Table 6.2 shows considerable heterogeneity in both business sector and firm size.

The results of our analysis are an important step in understanding systemic risk and its potential origins at the level of firms. They suggest that firms posing systemic risk have more heterogeneity than the focus on financial firms has led many to believe. Instead, our estimations demonstrate that many of the most central firms—and thereby firms posing the risk of largest aggregate losses following an idiosyncratic shock—are of varying sizes and in varying industries. Furthermore, focusing on those aggregate losses as a ratio of firm revenue revealed how some firms have a disproportionate impact on the economy through a multiplier effect borne out of network ties. The findings point to the value of further studies on the role of firm-level networks and their potential value to macroeconomic policy.

## Conclusions and Implications

---

This report contributes to a growing literature on the microeconomic origins of aggregate losses from systemic risk in the broad economy. Although existing literature focuses on the use of sector-level models, this report contributes a novel approach for assessing interfirm production networks to estimate the aggregate effects of idiosyncratic shocks to individual firms. This approach provides a path toward identifying central firms across diverse sectors within the U.S. economy, revealing areas of systemic risk in traditionally understudied areas.

Our estimation of interfirm network connections leveraged publicly available data on firm supplier-customer linkages. Although this data—from an SEC requirement on public companies to report key customers and suppliers—is available from a variety of financial data services, we extracted it directly from the SEC’s EDGAR reporting system. The data provided a network of observed connections between firms. The network is partial, but in terms of missing data on unobserved connections and flows across observed connections, the data set provided a rich network for study. Using network analysis on the observed connections, we gained insight into the structure of the network between firms, including the level of interconnections. Although many firms exist with few connections in a sparse network, some exist in heavily interconnected areas of the network. Of those heavily interconnected firms, we observed that many represent top financial firms (e.g., Bank of America, J.P. Morgan) while others represent top technology (e.g., Alphabet, Amazon, Apple, Cisco), telecommunications (e.g., AT&T), and health care (e.g., UnitedHealth Group, CVS

Health) firms. The results of the observed network suggest that systemically important firms can come from many sectors.

To understand the economic centrality of firms, we estimated the potential complete network configurations. This approach used inference to estimate likely unobserved network connections and optimization to calibrate networks to firm data, thereby estimating the flow of goods and services across estimated network linkages. Taking this approach resulted in expanding traditional input-output modeling from sector-level analysis to firm-level analysis, a contribution to computational economic modeling. It also provided new insights into the systemic importance of firms. Using estimated network connections to calculate the potential impact of idiosyncratic shocks, we found that few firms represent the potential for significant aggregate losses. However, of those that do, we found that many extend beyond the traditional sector of systemic concern (i.e., finance) and into diverse sectors from manufacturing, to healthcare, to telecommunications and to technology.

Our analysis of systemic risk in the broad economy highlights a potential shift in the landscape of systemic risk a decade after the 2008 financial crisis. Although that crisis revealed the systemic risk of financial sector firms (namely banks, which were thought to be both too big and too interconnected to fail), the economy has shifted since 2008. Many of the systemically important firms we uncovered (e.g., Amazon and Alphabet) have undergone extraordinary periods of growth during this decade with market power that reaches millions of customers with a variety of increasingly essential services. Although these firms' potential for systemic risk is different—that is, they might not exhibit the inherent risk of bank runs that financial firms do—our results underscore the need for new policy discussions surrounding emerging risks.

## Policy Implications

This report represents a step forward in filling a gap in the discussion surrounding the potential for contagion following firm-level shocks.

Although we found that financial-sector firms are some of the most interconnected in the economy, we also found that many of the most-connected firms actually exist in other sectors. Specifically, firms in both the technology and telecommunications sectors represent some of the most connected in observed production networks. When economic centrality is considered, we found that technology and telecommunications firms top our estimations of most central.

This observation should come as a surprise to few. In the years since 2008, U.S. firms and households have become increasingly reliant on the services of such large firms as Amazon, Comcast, Apple, and Alphabet, so that a shock to any individual firm could ripple across every sector. Focusing on Amazon, just one of its core businesses, Amazon Web Services (AWS), has grown from \$12 billion in 2016 annual sales to more than \$25 billion in annual sales in 2018, providing cloud internet services to millions of customers, such as households, firms, and government entities (Amazon, 2018). The potential for systemic effects of a sustained shock to just its AWS service is notable.

The centrality of technology and telecommunications firms is especially important given a changing landscape in global risk. In the decade since the financial crisis, the focus of many policymakers has shifted to cybersecurity. The 2016 cyberattack on Dyn, a relatively small firm providing a key domain name system, did more damage than just shock to the single firm; it knocked out services to all of its customers, representing numerous large online firms in diverse industries, including Netflix, Amazon, and the *New York Times* (Meyer and Lafrance, 2016). Although the Dyn incident could be an isolated one, it could also be a warning, similar to that signaled by the failure of Long Term Capital Management, of an emerging source of systemic risk. The potential impacts of shocks to AWS could have more-significant aggregate impacts. Furthermore, shifts in global risk are not contained to cybersecurity: Such challenges as severe weather might lead to new concentrations of risk that could propagate across the production networks of many sectors and not remain contained within the affected geography.

The result is a need for a renewed policy discussion surrounding the systemic risk of firms across the broader economy. Our estimations

are not yet conclusive but rather point to the need for further investigation by relevant governmental agencies and researchers. Components of the Department for Homeland Security, for example, should seek to understand which firms are vulnerable in the presence of cyberthreats and which firms are vulnerable in the presence of extreme climate events. This effort might leverage those that have started since 2008; in addition to its study of financial market stability, the U.S. Treasury Department's Office of Financial Research has already begun studying the intersection of cybersecurity risks and market stability (U.S. Department of the Treasury, Office of Financial Research, 2017). In a manner parallel to the policy debate following the 2008 financial crisis, policymakers might need to begin answering the question of whether other firms, particularly in tech and telecommunications, have become too big and too interconnected and, if so, they might need to identify potential mitigations through exercises analogous to stress tests used for systemically important banks.

## Limitations and Future Work

The contributions in this report are an important step for understanding firm-level production networks, the microfoundations of aggregate risk, and the systemic risk posed by individual firms across the broad economy. They are also a first step with many notable limitations. To start, our methodology and findings depended on observed partial network connections. Without ground truth on actual flows between firms, this report and others will be limited in their ability to precisely assess the contributions of individual firms to systemic risk. As a result, although our work highlights the ability for firms across the economy to contribute to systemic risk, it would not suffice as a prescriptive tool for policy changes surrounding a single individual firm.

Furthermore, our methodological contributions aimed to infer missing data from partially observed production networks which, of course, pose limitations. The estimation approaches presented in this report are imprecise. They could, and should, be further refined for higher levels of predictive accuracy. We point to machine-learning

approaches as a likely tool for enhanced estimation procedures. Furthermore, although we presented an effective tool for calibrating estimated production networks by computational optimization, future work could enhance the robustness of this approach, and its computational efficiency.

Above all, this work highlights a direction for further research: Our results are only a show of principle. Further research on predicting unobserved firm connections, and on the strength of their economic ties, is necessary for bringing this research to its full maturity. Future work should leverage interdisciplinary methods, as advances in the fields of machine learning and scalable analytics can lead to robust estimation approaches. The value of firm-level analysis might prove a powerful tool in the field of computational economics for evaluating the potential aggregate losses associated with isolated incidents. Although we give particular attention to macroeconomics, this approach can be taken by others exploring the potential impact of environmental risks, critical infrastructure, and even intentional threats.





## References

---

- Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The Network Origins of Aggregate Fluctuations," *Econometrica*, Vol. 80, No. 5, September 2012, pp. 1977–2016.
- Acemoglu, Daron, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "Systemic Risk and Stability in Financial Networks," *American Economic Review*, Vol. 105, No. 2, February 2015, pp. 564–608.
- Allen, Franklin, and Douglas Gale, "Financial Contagion," *Journal of Political Economy*, Vol. 108, No. 1, February 2000, pp. 1–33.
- Amazon, "2018 Annual Report," webpage, 2018. As of November 4, 2019: <https://ir.aboutamazon.com/annual-reports>
- Arata, Yoshiyuki, *Bankruptcy Propagation on a Customer-Supplier Network: An Empirical Analysis in Japan*, Research Institute of Economy, Trade and Industry, Discussion Paper 18-E-040, 2018.
- Barrot, Jean-Noël, and Julien Sauvagnat, "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks," *Quarterly Journal of Economics*, Vol. 131, No. 3, August 2016, pp. 1543–1592.
- Bigio, Saki, and Jennifer La'O, *Financial Frictions in Production Networks*, National Bureau of Economic Research, Working Paper Series, No. 22212, 2016.
- Blanchard, Olivier J., David Romer, A. Michael Spence, and Joseph E. Stiglitz, eds., *In the Wake of the Crisis: Leading Economists Reassess Economic Policy*, Vol. 1, Cambridge, Mass.: MIT Press, 2012.
- Bonacich, Phillip, "Power and Centrality: A Family of Measures," *American Journal of Sociology*, Vol. 92, No. 5, March 1987, pp. 1170–1182. As of
- Bonacich, Phillip, and Paulette Lloyd, "Eigenvector-Like Measures of Centrality for Asymmetric Relations," *Social Networks*, Vol. 23, No. 3, July 2001, pp. 191–201.
- Carvalho, Vasco M., "From Micro to Macro via Production Networks," *Journal of Economic Perspectives*, Vol. 28, No. 4, Fall 2014, pp. 23–48.

Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alireza Tahbaz-Salehi, *Supply Chain Disruptions: Evidence from the Great East Japan Earthquake*, Columbia Business School Research Paper, No. 17–5, December 2016.

Carvalho, Vasco M., and Alireza Tahbaz-Salehi, “Production Networks: A Primer,” *Annual Review of Economics*, Vol. 11, No. 1, 2019, pp. 635–663.

Diamond, Douglas W., and Philip H. Dybvig, “Bank Runs, Deposit Insurance, and Liquidity,” *Journal of Political Economy*, Vol. 91, No. 3, June 1983, pp. 401–419.

Dornbusch, Rudiger, Yung Chul Park, and Stijn Claessens, “Contagion: Understanding How It Spreads,” *World Bank Research Observer*, Vol. 15, No. 2, August 2000, pp. 177–197.

Eisenberg, Larry, and Thomas H. Noe, “Systemic Risk in Financial Systems,” *Management Science*, Vol. 47, No. 2, February 2001, pp. 236–249.

Giesecke, Kay, and Stefan Weber, “Credit Contagion and Aggregate Losses,” *Journal of Economic Dynamics and Control*, Vol. 30, No. 5, May 2006, pp. 741–767.

Glasserman, Paul, and H. Peyton Young, “How Likely is Contagion in Financial Networks?” *Journal of Banking and Finance*, Vol. 50, January 2015, pp. 383–399.

Goolsbee, Austan D., and Alan B. Krueger, “A Retrospective Look at Rescuing and Restructuring General Motors and Chrysler,” *Journal of Economic Perspectives*, Vol. 29, No. 2, Spring 2015, pp. 3–24.

Kiyotaki, Nobuhiro, and John Moore, *Credit Cycles*, National Bureau of Economic Research Working Paper Series, No. 5083, April 1995.

Kodres, Laura E., and Matthew Pritsker, “A Rational Expectations Model of Financial Contagion,” *Journal of Finance*, Vol. 57, No. 2, April 2002, pp. 769–799.

Leontief, Wassily, ed., *Input-Output Economics*, New York: Oxford University Press, 1986.

Long, John B., Jr., and Charles I. Plosser, “Real Business Cycles,” *Journal of Political Economy*, Vol. 91, No. 1, February 1983, pp. 39–69.

MacKenzie, Cameron A., Kash Barker, and Joost R. Santos, “Modeling a Severe Supply Chain Disruption and Post-Disaster Decision Making with Application to the Japanese Earthquake and Tsunami,” *IIE Transactions*, Vol. 46, No. 12, December 2014, pp. 1243–1260.

Meyer, Robinson, and Adrienne Lafrance, “When the Entire Internet Seems to Break at Once: The Easiest Way to Take Down the Web Is to Attack People’s Access to It,” *The Atlantic*, October 21, 2016. As of November 4, 2019: <https://www.theatlantic.com/technology/archive/2016/10/when-the-entire-internet-seems-to-break-at-once/504956/>

Public Law 111-203, Dodd-Frank Wall Street Reform and Consumer Protection Act, July 21, 2010.

Ricci, Vito, “Fitting Distributions with R,” release 0.4, February 21, 2005. As of November 8, 2019:

<https://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf>

U.S. Department of the Treasury, Office of Financial Research, “Cybersecurity and Financial Stability: Risks and Resilience,” *Viewpoint 17-01*, February 2017.

Wu, Jing, and John R. Birge, “Supply Chain Network Structure and Firm Returns,” January 2014.



In the years following the 2008 financial crisis, significant attention was paid to systemic risk within heavily interconnected financial networks. Academic discussions gave rise to a policy debate about whether major banks had become both too big and too interconnected to fail. However, despite the focus on systemic risk—the risk of market collapse resulting from firm-level risks—within the financial sector, little attention was paid to systemic risks in the economy at large. Using network analysis, statistical inference, and network calibration, the authors of this report provide a new approach to modeling the economy at the firm level that expands on the traditional sector-level input-output modeling by estimating firm-level input-output flows. The result allows one to use traditional input-output modeling to estimate the size of potential idiosyncratic shocks and to use economically weighted measures of centrality to reveal systemically important firms. The approach is a contribution to the growing literature on the microfoundations of economic risk, with the potential for use across a wide range of applications from financial stability to natural disasters.

\$15.00

ISBN-10 1-9774-0390-5  
ISBN-13 978-1-9774-0390-2



[www.rand.org](http://www.rand.org)